Bayesian Semiparametric GARCH Models

Xibin (Bill) Zhang and Maxwell L. King

Department of Econometrics and Business Statistics
Faculty of Business and Economics
MONASH University

xibin.zhang@monash.edu

Quantitative Methods in Finance, 14–17 December 2011

Outline

Motivation

- The GARCH model of Bollerslev (1986) has been useful in modelling volatilities of asset returns.
- Let $\boldsymbol{y} = (y_1, \cdots, y_n)'$ denote a vector of n observations of an asset return. A GARCH(1,1) model is often specified as

$$\begin{aligned} y_t &= \sigma_t \, \varepsilon_t, \\ \sigma_t^2 &= \omega + \alpha y_{t-1}^2 + \beta \sigma_{t-1}^2, \end{aligned} \tag{1}$$

where ε_t, for $t = 1, 2, \cdots, n$, are independent.

- The assumption of conditional normality of ε_t has contributed to early successes of GARCH models.
- For example, the QMLEs of parameters are consistent when the first two moments of y_t are correctly specified.

Problems

- However, enough evidence has shown that it is possible to reject the assumption of conditional normality.
- This has motivated the investigation of other specifications of the conditional distribution of errors in GARCH models, such as the Student t and other heavy-tailed distributions.
- Any assumption on the analytical form of the error density is only an approximation to the unknown true error density.
- Moreover, it is very important to investigate the distribution of response, which is implied by the error density.
- Some investigations have been focused on parameter estimation for (G)ARCH models without assumptions on the analytical form of the error density.

A Semiparametric Approach

- Engle and González-Rivera (1991) proposed a semiparametric GARCH model without assumptions on the analytical form of error density.
- The error density was estimated nonparametrically based on residuals, which were obtained by applying either the QMLE (assuming conditional normality) or OLS.
- The parameters of the semiparametric GARCH model were estimated by maximising the log-likelihood constructed through the estimated error density.
- Their Monte Carlo study showed that this semiparametric approach could improve the efficiency of parameter estimates up to 50% against QMLEs obtained under conditional normality.

Limitations and Our Aims

- Their likelihood is affected by initial parameter estimates.
- Their semiparametric estimation uses the data twice because residuals have to be pre-fitted to construct likelihood.
- Their derived semiparametric estimates of parameters would not be used again to improve the error density estimator.
- We propose to approximate the true error density by a mixture of n normal densities, which have a common variance and individual means at the errors. This mixture density has the form of kernel density estimator of the errors.
- We treat the re-parameterised bandwidth as an additional parameter and investigate the likelihood and posterior under this mixture density of errors.

A location-mixture of n Gaussian densities

- We assume that the unknown density of ε_t denoted as $f(\varepsilon_t)$, is approximated by a location-mixture density:

$$f(\varepsilon_t; h) = \frac{1}{n}\sum_{i=1}^{n} \frac{1}{h}\phi\left(\frac{\varepsilon_t - \varepsilon_i}{h}\right), \tag{2}$$

for $t = 1, 2, \ldots, n$, where $\phi(\cdot)$ is the Gaussian PDF, and the component Gaussian densities have a common variance h^2 and different means at ε_i, for $i = 1, 2, \ldots, n$.

- From the view of kernel smoothing, this mixture error density is a kernel density estimator of errors with h the bandwidth. Therefore, we call $f(\varepsilon; h)$ the mixture (or kernel-form) error density, where h is called either the standard deviation or bandwidth, which determines the performance of $f(\varepsilon_t; h)$.

Our contribution

- We propose to approximate the unknown error density by the kernel-form error density, based on which we are able to construct the likelihood.
- We choose prior densities for the two types of parameters and obtain the posterior of all parameters. Bayesian sampling is then used to sample these parameters.
- We use this semiparametric GARCH model to compute value-at-risk (VaR).

Investigations related our work

- The validity of this mixture density as a density of the regression errors was investigated by Yuan and de Gooijer (2007) in a class of nonlinear regression models.
- Our work is related to adaptive estimation discussed by Linton (1993) and Drost and Klaassen (1997) for (G)ARCH models. A conclusion from their work is the parameters are approximately adaptively estimable.
- In all these studies, bandwidth was chosen based on pre-fitted residuals, which were used as proxies of errors.
- Bayesian semiparametric estimation was discussed by Koop (1992) for ARCH models, where the quasi likelihood was set up through a sequence of complicated polynomials.

Kernel-form conditional density of errors

- Consider the GARCH(1,1) model

$$\begin{aligned} y_t &= \sigma_t \, \varepsilon_t, \\ \sigma_t^2 &= \omega + \alpha y_{t-1}^2 + \beta \sigma_{t-1}^2, \end{aligned} \tag{3}$$

where $\omega > 0$, $\alpha \geq 0$, $\beta \geq 0$ and $\alpha + \beta < 1$.

- Strictly speaking, conditional on information available at $t-1$ denoted as I_{t-1}, the density of ε_t denoted as $f(\varepsilon_t)$, is unknown.
- When $f(\varepsilon)$ is assumed to be known, the likelihood is

$$\ell_0(\boldsymbol{y}|\omega, \alpha, \beta) = \prod_{t=1}^{n} \frac{1}{\sigma_t} f\left(\frac{y_t}{\sigma_t}\right).$$

- $f(\varepsilon)$ could be the Gaussian, Student t, and a mixture of several Gaussian densities.

A location-mixture density of n Gaussian densities

- We propose a location-mixture density as the error density:

$$f(\varepsilon; h) = \frac{1}{n} \sum_{i=1}^{n} \frac{1}{h} \phi\left(\frac{\varepsilon - \varepsilon_i}{h}\right),$$

where the component Gaussian densities have a common variance h^2 and individual means at individual errors.

- In some investigations on asymptotic properties of parameter estimators, bandwidth depends on sample size n and approaches zero as $n \to \infty$.

- We re-parameterise h as $\tau n^{-1/5}$, where $n^{-1/5}$ is the optimal convergence rate under AMISE. Hereafter, bandwidth is denoted as $h_n = \tau n^{-1/5}$.

Benefit of the mixture error density

- $f(\varepsilon; h_n)$ has the form of kernel density estimator of errors.

- In terms of kernel density estimation based on directly observed data, Silverman (1978) proved that a kernel density estimator approaches the underlying true density as $n \to \infty$.

- Therefore, it is reasonable to expect that $f(\varepsilon; h_n)$ approaches $f(\varepsilon)$ as the sample size n increases.

- $f(\varepsilon_t; h_n)$ differs from the kernel density estimator of residuals calculated through pre-estimated parameters. This mixture density is defined conditional on model parameters:

$$f(\varepsilon_t; h_n) = \frac{1}{n} \sum_{i=1}^{n} \frac{1}{h_n} \phi\left(\frac{\varepsilon_t - y_i/\sigma_i}{h_n}\right), \tag{4}$$

where $\sigma_i^2 = \omega + \alpha y_{i-1}^2 + \beta \sigma_{i-1}^2$, for $i = 1, 2, \ldots, n$.

Benefit of the mixture error density (2)

- To construct likelihood, we use the leave-one-out:

$$f(\varepsilon_t|\boldsymbol{\varepsilon}_{(t)}; h_n) = \frac{1}{n-1}\sum_{\substack{i=1\\i\neq t}}^{n}\frac{1}{h_n}\phi\left(\frac{\varepsilon_t-\varepsilon_i}{h_n}\right). \quad (5)$$

- The density of y_t is estimated by

$$f_Y\left(y_t|\boldsymbol{y}_{(t)}; \theta\right) = \frac{1}{(n-1)\sigma_t}\sum_{\substack{i=1\\i\neq t}}^{n}\frac{1}{h_n}\,\phi\left(\frac{y_t/\sigma_t-y_i/\sigma_i}{h_n}\right). \quad (6)$$

- In the density function of y_t given by (6), h_n and σ_t always appear in the form of the product of the two:

$$h_n^2\sigma_t^2 = h_n^2\omega + h_n^2\alpha y_{t-1}^2 + \beta h_n^2\sigma_{t-1}^2. \quad (7)$$

Likelihood

- h_n^2 and ω, as well as h_n^2 and α, cannot be separately identified. If ω (or α) is assumed to be a known constant, all the other parameters can be separately identified.
- In adaptive estimation for ARCH models, ω was restricted to be zero by Linton (1993) and one by Drost and Klaassen (1997).
- In light of the fact that the unconditional variance of y_t is $\omega/(1-\alpha-\beta)$, we assume that $\omega = (1-\alpha-\beta)s_y^2$.
- When the return series is pre-standardised, the value of ω would be assumed to be $(1-\alpha-\beta)$, which is the same as what Engle and Gonzlez-Rivera (1991) assumed for ω in their GARCH model.

Likelihood and Choices of Priors

- The parameter vector is $\theta = \left(\sigma_0^2, \alpha, \beta, \tau^2\right)'$, and the restrictions are: $0 \leq \alpha < 1$, $0 < \beta < 1$ and $0 < \alpha + \beta < 1$.
- The likelihood of $\boldsymbol{y} = (y_1, y_2, \ldots, n)'$, for given θ, is

$$\ell(\boldsymbol{y}|\theta) = \prod_{t=1}^{n} \left\{ \frac{1}{(n-1)\sigma_t} \sum_{\substack{i=1 \\ i \neq t}}^{n} \frac{1}{h_n} \, \phi\left(\frac{y_t/\sigma_t - y_i/\sigma_i}{h_n} \right) \right\}. \tag{8}$$

- Conditional on model parameters, this likelihood function is the one used by the likelihood cross-validation in choosing bandwidth for the kernel density estimator of standardised y_i.
- The prior of α is the uniform density on $(0, 1)$, and the prior of β is the uniform density on $(0, 1-\alpha)$.

Prior Choices and Posterior

- As $\left(\tau n^{-1/5}\right)^2$ is the variance of component Gaussian densities in the mixture, we assume $\left(\tau n^{-1/5}\right)^2$ follows an inverse Gamma distribution denoted as $\mathrm{IG}(a_\tau, b_\tau)$. Therefore, the prior of τ^2 is

$$p(\tau^2) = \frac{b_\tau^{a_\tau}}{\Gamma(a_\tau)} \left(\frac{1}{\tau^2 n^{-2/5}} \right)^{a_\tau+1} \exp\left\{ -\frac{b_\tau}{\tau^2 n^{-2/5}} \right\} n^{-2/5}.$$

- The prior of σ_0^2 is assumed to be either the log normal density with mean zero and variance one or the density of IG(1,0.05).
- The joint prior of θ denoted as $p(\theta)$, is the product of the marginal priors of α, β, τ^2 and σ_0^2.
- The posterior of θ for given $\boldsymbol{y}$ is proportional to the product of the joint prior of θ and the likelihood of $\boldsymbol{y}$ for given θ:

$$\pi(\theta|\boldsymbol{y}) \propto p(\theta) \times \ell(\boldsymbol{y}|\theta).$$

Data, models and results

- We used the random-walk Metropolis algorithm to sample parameters of the GARCH(1,1) model of daily returns of the S&P 500 index.
- The sample period is from 03/01/2007 to 30/06/2011, and the sample size is $n = 1131$
- We estimated the parameters in the semiparametric GARCH(1,1) model and t-GARCH(1,1) model. Results are presented in Tables 1 and 2.
- The prior of the degrees-of-freedom parameter ν is $N(10, 5^2)$ truncated at 3, and the prior of ω is $U(0, 1)$.

Table 1: Results from the semiparametric GARCH(1,1) model.

Parameters	Mean	95% Bayesian credible interval	Batch-mean SD	Standard deviation	SIF
σ_0^2	0.496103	$(0.0875, 1.5504)$	0.011461	0.390368	8.62
α	0.082482	$(0.0593, 0.1103)$	0.000789	0.013433	34.51
β	0.892831	$(0.8557, 0.9241)$	0.001118	0.018271	37.45
τ	0.793211	$(0.5247, 1.0873)$	0.006063	0.142889	18.00
log marginal likelihood	-1839.72				

Table 2: Results from the t-GARCH(1,1) model.

Parameters	Mean	95% Bayesian credible interval	Batch-mean SD	Standard deviation	SIF
σ_0^2	0.335206	$(0.0789, 0.8520)$	0.005601	0.217653	6.62
ω	0.015697	$(0.0040, 0.0240)$	0.000472	0.006869	47.23
α	0.073472	$(0.0466, 0.1003)$	0.000611	0.013699	19.92
β	0.890709	$(0.8492, 0.9210)$	0.001030	0.018130	32.29
ν	6.807922	$(3.8381, 7.6489)$	0.063646	1.332099	22.83
log marginal likelihood	-1855.30				

Convergence performance of our sampler

- The burn-in period contains 1000 draws, and the following 10,000 draws were recorded.
- We computed the batch-mean standard deviation and simulation inefficiency factor (SIF) to monitor the convergence.
- The SIF is approximately interpreted as the number of draws needed to derive independent draws.
- For example, a SIF value of 20 means that approximately, we should retain 1 draw for every 20 draws to obtain independent draws in this sampling procedure.
- All simulated chains under the mixture error density have achieved very reasonable convergence.
- The marginal likelihood derived under the mixture error density is obviously larger than that derived under the t errors.

Bayes factor for model comparison

- Bayes factor is a ratio of the marginal likelihoods derived under a model of interest and its competing model.
- Let θ denote the parameter vector under model $\mathcal{A}$. The marginal likelihood under model $\mathcal{A}$ is (Chib, 1995)

$$m_{\mathcal{A}}(\boldsymbol{y}) = \frac{\ell_{\mathcal{A}}(\boldsymbol{y}|\theta)p_{\mathcal{A}}(\theta)}{\pi_{\mathcal{A}}(\theta|\boldsymbol{y})}.$$

$\ell_{\mathcal{A}}(\boldsymbol{y}|\theta)$ and $p_{\mathcal{A}}(\theta)$ are likelihood and prior under model $\mathcal{A}$.

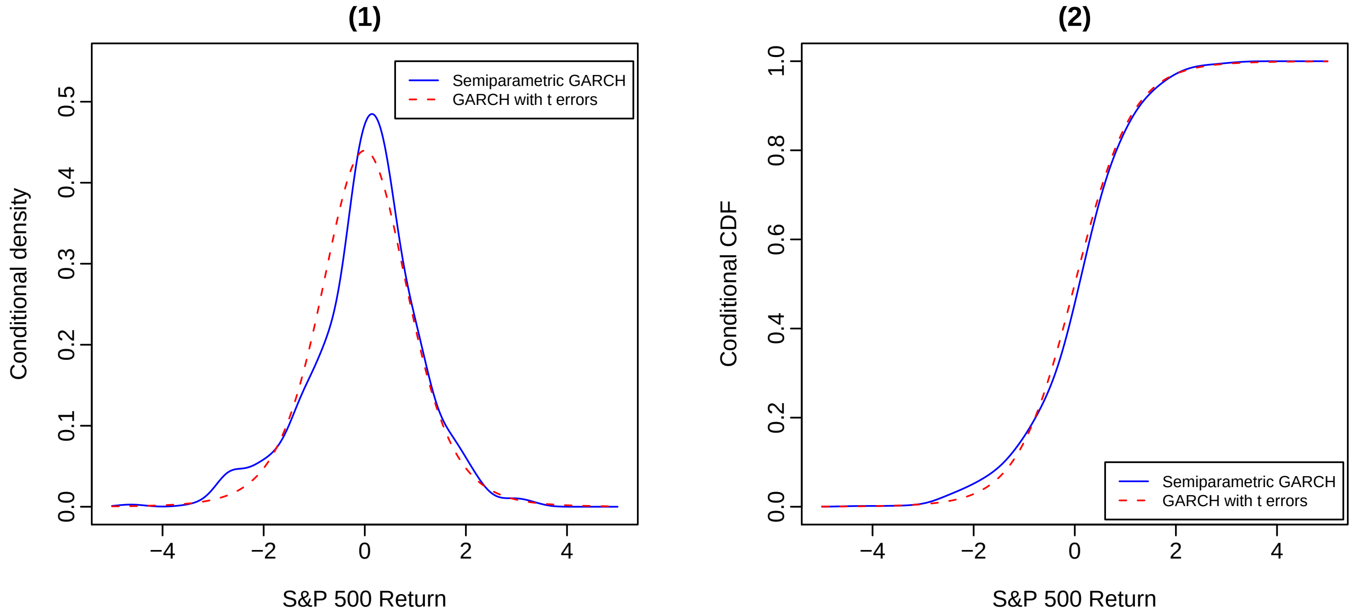
- The Bayes factor of model $\mathcal{A}$ against model $\mathcal{B}$ is

$$\text{BF} = \frac{m_{\mathcal{A}}(\boldsymbol{y})}{m_{\mathcal{B}}(\boldsymbol{y})}.$$

- $3 < \text{BF} \leq 20$: $\mathcal{A}$ is favored against $\mathcal{B}$ with positive evidence.
- $20 < \text{BF} \leq 150$: $\mathcal{A}$ is favored against $\mathcal{B}$ with strong evidence.
- $\text{BF} > 150$: $\mathcal{A}$ is favored against $\mathcal{B}$ with very strong evidence.

Density forecast of the one-step out-of-sample S&P 500 return

Figure 1: The estimated densities and CDFs of the one-step out-of-sample return: (1) Conditional density of y_{n+1}; and (2) conditional CDF of y_{n+1}.



Conditional VaR

- The VaRs under the semiparametric and t GARCH models are \$2.0324 and \$1.6643 for a \$100 investment on S&P 500.
- Therefore, in comparison to the semiparametric GARCH model, the t-GARCH tends to underestimate VaR.

Motivation for localised bandwidths

- When the true error density has sufficient long tails, the leave-one-out kernel density estimator with its bandwidth selected under the Kullback-Leibler criterion, is likely to overestimate the tails density.
- One may argue that this phenomenon is likely to be caused by the use of a global bandwidth. A remedy to this problem in that situation is to use variable bandwidths or localized bandwidths.
- Small bandwidths should be assigned to the observations in the high-density region and larger bandwidths should be assigned to those in the low-density region.

Localised bandwidths

- We assume the underlying true error density is unimodal. Large absolute errors should be assigned relatively large bandwidths, while small absolute errors should be assigned relatively small bandwidths.
- We propose the following error density estimator:

$$f_a\left(\varepsilon_t;\tau,\tau_\varepsilon\right)=\frac{1}{n-1}\sum_{\substack{i=1\\i\neq t}}^{n}\frac{1}{\tau n^{-1/5}\left(1+\tau_\varepsilon\left|\varepsilon_i\right|\right)}\,\phi\left(\frac{\varepsilon_t-\varepsilon_i}{\tau n^{-1/5}\left(1+\tau_\varepsilon\left|\varepsilon_i\right|\right)}\right),$$

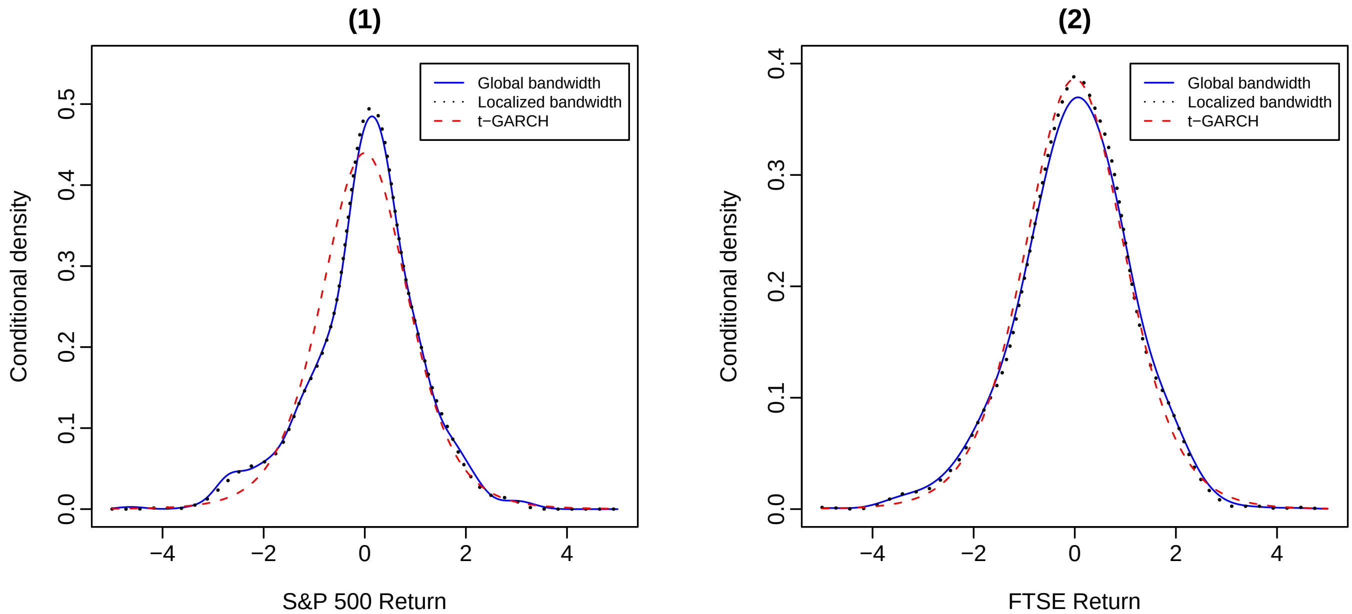
where $\tau n^{-1/5}\left(1+\tau_\varepsilon\left|\varepsilon_i\right|\right)$ is the bandwidth assigned to ε_i, and the vector of parameters is now $\theta_a=\left(\sigma_0^2,\alpha,\beta,\tau,\tau_\varepsilon\right)'$.

Bayesian estimate

- The prior of τ_ε is the uniform density on $(0, 1)$.
- The parameter estimates are $\alpha = 0.093154$, $\beta = 0.893324$, $\tau = 0.763784$, $\tau_\varepsilon = 0.635660$, and $\sigma_0^2 = 0.427865$.
- The sampler has converged very well.
- The log marginal likelihood under localised bandwidths (global bandwidth) is -1835.67 (-1839.72).
- The Byayes factor of the use of localised bandwidths against the use of global bandwidth is exp(4.05), and the former is favored against the latter with strong evidence.
- The use of localised bandwidths has increased the competitiveness of the semiparametric GARCH model.

Density forecast of the one-step out-of-sample return

Figure 1: The estimated densities of the one-step out-of-sample return through localised bandwidths: (1) S&P 500 return; and (2) FTSE return.



Application to other index returns

Table 3(1): Results from semiparametric GARCH (global bandwidth)

	Nasdaq	NYSE	DJIA	FTSE	DAX	AORD	Nikkei
σ_0^2	0.737396	0.675171	0.452820	0.672453	0.757102	1.021040	0.841057
α	0.098014	0.086032	0.094623	0.108520	0.144759	0.137622	0.137127
β	0.887498	0.892873	0.883619	0.880093	0.854487	0.851735	0.844828
τ	0.977918	1.017463	1.241229	1.360150	1.082082	1.228259	1.050393
VaR	2.3327	2.1847	1.8587	2.0247	2.2087	1.9937	2.0937
LML	-1959.64	-1924.85	-1748.03	-1880.29	-1955.95	-1791.04	-2013.62

Table 3(2): Results from the t-GARCH(1,1) model.

	Nasdaq	NYSE	DJIA	FTSE	DAX	AORD	Nikkei
σ_0^2	0.672198	0.471396	0.312151	0.576110	0.638668	0.946476	0.816669
ω	0.024144	0.021031	0.011339	0.030545	0.028463	0.031719	0.064810
α	0.071868	0.073784	0.074238	0.084517	0.067160	0.097714	0.112387
β	0.892783	0.891033	0.891644	0.875496	0.895703	0.860881	0.837827
ν	7.619080	7.530589	6.697563	9.158895	8.110169	10.632002	9.962462
VaR	2.0407	1.8027	1.5467	1.8387	1.9207	1.7917	1.9037
LML	-1973.36	-1936.98	-1750.14	-1874.35	-1950.56	-1788.66	-2023.34

Application to other index returns

Table 4: Results from semiparametric GARCH (localised bandwidth).

	S&P 500	Nasdaq	NYSE	DJIA	FTSE	DAX	AORD	Nikkei
σ_0^2	0.4279	0.7510	0.6235	0.3831	0.6802	0.7610	1.0060	0.8383
	(6.20)	(3.81)	(6.44)	(4.92)	(7.40)	(9.15)	(8.06)	(8.28)
α	0.0932	0.0958	0.0940	0.1085	0.1047	0.0988	0.1190	0.1460
	(14.60)	(14.63)	(13.95)	(10.61)	(8.19)	(42.14)	(15.11)	(13.72)
β	0.8933	0.8917	0.8915	0.8795	0.8813	0.8917	0.8668	0.8366
	(15.47)	(17.07)	(16.13)	(13.43)	(8.80)	(23.14)	(17.35)	(15.99)
τ	0.7638	0.8094	0.7470	0.8073	0.8364	0.8423	0.7797	0.7389
	(12.19)	(13.15)	(13.71)	(12.77)	(22.95)	(27.36)	(23.98)	(20.18)
τ_ε	0.6357	0.4174	0.6540	0.7617	0.5283	0.6878	0.5305	0.4276
	(28.08)	(14.89)	(18.41)	(24.20)	(23.53)	(29.69)	(25.97)	(21.58)
VaR	1.9778	2.3147	2.1587	1.8097	1.9687	2.1207	1.9227	2.0717
LML	-1835.67	-1958.63	-1918.19	-1733.59	-1876.13	-1949.46	-1787.70	-2012.25

Table 5: A summary of marginal likelihoods.

	S&P 500	Nasdaq	NYSE	DJIA	FTSE	DAX	AORD	Nikkei
t	-1855.30	-1973.36	-1936.98	-1750.14	-1874.35	-1950.56	-1788.66	-2023.34
Global	-1839.72	-1959.64	-1924.85	-1748.03	-1880.29	-1955.95	-1791.04	-2013.62
Localised	-1835.67	-1958.63	-1918.19	-1733.59	-1876.13	-1949.46	-1787.70	-2012.25

Empirical findings

- The semiparametric GARCH with localised bandwidths is favored against the t-GARCH for S&P 500, Nasdaq, NYSE, DJIA, and Nikkei 225 indices. The t-GARCH is favored against the former for FTSE.
- The use of localised bandwidths increases the competitiveness against its competitor, the t-GARCH model. This is evidenced by an increased marginal likelihood for each index.
- The use of localized bandwidths slightly reduces the VaR compared to the use of a global bandwidth, but the relative change is between 0.77% to 3.98%.
- The t-GARCH model underestimates VaR in comparison to the semiparametric GARCH model.

Conclusion

- We approximate unknown error density by a location-mixture density of n normal densities for GARCH models.
- We derived the likelihood and posterior for all parameters, and Bayesian sampling is conducted for estmation.
- The mixture error density is used to forecast the density of the one-day out-of-sample return, which is used for estimating VaR.
- The use of localised bandwidths increases the competitiveness the semiparametric GARCH against the t-GARCH.
- The semiparametric GARCH is favored against the t-GARCH for five out of eight indices.
- The t-GARCH underestimates VaR compared to the semiparametric GARCH.